DOCUMENT RESUME

ED 428 214 CE 078 247

TITLE	Measurement and Research Tools.
PUB DATE	1997-03-00
NOTE	27p.; In: Academy of Human Resource Development (AHRD) Conference Proceedings (Atlanta, Georgia, March 6-9, 1997); see CE 078 221.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Adult Education; Cognitive Style; *Computer Software; Computer Uses in Education; *Data Analysis; Evaluation Methods; Human Resources; *Labor Force Development; *Learning Modalities; *Multiple Regression Analysis; Teaching Methods; Testing; Tests
IDENTIFIERS	*Loyalty

ABSTRACT

This document contains four papers from a symposium on measurement and research tools for human resource development (HRD). "The 'Best Fit' Training: Measure Employee Learning Style Strengths" (Daniel L. Parry) discusses a study of the physiological aspect of sensory intake known as modality, more specifically, modality as measured by the Multi-Modal Paired Associates Learning Test versions II and III. "Measuring Customer Loyalty" (Karen A. Brennan) reports on a pilot study that identified criteria to define loyal and non-loyal customers: customer-generated revenue, customer satisfaction, willingness to refer, and time interval since last purchase. "Textual Analysis with Neural Networks as an Alternative Assessment Tool for Evaluating Responses to Open-Ended Questions" (William Venable, F. George McDuffee) discusses experiences in using neural network-based software to evaluate responses to open-ended questions on survey instruments, such as strengths and limitations of the software package used, problems encountered, "work arounds" for these problems, and suggestions for incorporating this or a similar package for free form text analysis. "The Impact of Influential Observations on Multiple Regression Analysis in HRD Research" (Reid A. Bates, Elwood F. Horton III, Michael F. Burnett) illustrates how a small number of observations can significantly influence results of regression analysis and outlines procedures for identifying influential observations. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 428 214

AHRD

ACADEMY OF HUMAN RESOURCE DEVELOPMENT

1997 Conference Proceedings

March 6-9, 1997
Atlanta, GA

25 **Measurement and Research Tools**

Chair: Wim J. Nijhof, University of Twente

The "Best Fit" Training: Measure Employee Learning Style Strengths
Daniel L. Parry, University of Louisville

Measuring Customer Loyalty
Karen A. Brennan, Zenger Miller

Textual Analysis with Neural Networks as an Alternative Assessment Tool for Evaluating Responses to Open Ended Questions
William Venable, Oklahoma State University
F. George McDuffee, Oklahoma State University

The Impact of Influential Observations on Multiple Regression Analysis in HRD Research
Reid A. Bates, Louisiana State University
Elwood F. Holton, III, Louisiana State University
Michael F. Burnett, Louisiana State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

DParry GMcDuffee
KBrennan EHolton

2A

CE078247

BEST COPY AVAILABLE

For "Best Fit" Training; Measure Employee Learning Style Strengths

Daniel L. Parry
University of Louisville

Employees must learn new skills to keep up with the changing technical nature of jobs. Since each employee receives information differently, instructors may find testing an employee's learning style modality to help the outcome of training. One tool that tests an individual's learning modality is the Multi-Modal Paired Associates Learning Test - III. The MMPALT-III may give practitioners an alternative to self-report instruments now in use..

Most human resource development (HRD) professionals would agree that jobs are changing, even without the benefit of hard data to support the statement. Computers, cellular phones, and fax machines are much more common than a short ten years ago. Similar changes have occurred in industry. Robotic manufacturing, computer aided design, e-mail, and video conferences are commonly found in today's business environment. It would be safe to say the skills that served an employee well in the past may be insufficient as job changes occur.

Research on the work force has shown workers in many areas have limited abilities. Jenkins and Kersch (1994) conducted a literacy profile with 1600 randomly selected subjects in the state of Ohio. The make-up of the subjects mirrored the working population of the Mid-West United States. A full 38% of the workers tested were functioning at the two lowest literacy levels. That is, those individuals had low problem solving skills and were able to complete tasks that involved only a single operation. Additionally, a report by the National Institute for Literacy (1990) stated:

> High wage jobs for people with low skills are on the decline. Increasingly, the only low-skill jobs will also be those that pay low wages. And greater job mobility is likely to mean people with lower skills will be out of work longer (p. 2).

If we accept that a significant number of workers have low literacy skills (and that only low paying jobs or lengthy unemployment await these workers) then we must also accept that for these workers to become part of a more modern work force, they need to be trained to improve their skills.

In order to make workers more viable in the modern work force, it is essential that some type of effective training take place. In the training of adults, learning styles have long been considered critical to the learning process (Wislock, 1993). The difficulty has been in defining learning styles. Cornett (1983) stated there are three broad descriptors of learning style:

Cognitive: how individuals decode, encode, process, and store information.

Affective: emotional and personality aspects of learning style.

Physiological: environmental factors and sensory intake.

Instructors may not be able to control the cognitive or affective aspects of learning, but can be aware of and control the physiological. Steps can be taken to ensure proper lighting, comfortable temperature, positive room decor and seating arrangements to address the environment. Instructors can also target training to the individual trainee's dominant mode of sensory intake to enhance training. Individuals who work in their dominant style(s) experience more success and feel a sense of calm and control (Grun, 1986), (James & Galbraeth, 1985).

This study focuses on the physiological aspect of sensory intake known as modality. More specifically, modality as measured by the Multi-Modal Paired Associates Learning Test versions II and III (MMPALT-II or MMPALT-III). The MMPALT-III is the current version of an objective test that uses paired associates testing methods to determine an individuals relative

©Copyright, Daniel L. Parry, 1997

modality strength. Modality is defined as the use of one of the senses to receive information and may be broken into one of seven types:

Print: Reading text only.
Aural: Hearing information.
Interactive: Discussion used in learning.
Visual: Associating pictures and text.
Haptic: Using whole hand touch.
Kinesthetic: Whole body movement.
Olfactory: Sensing odors.

Trainers have attempted to determine employees modalities through the use of self-report instruments and unfortunately these measure an individual's preference not strength. Studies by Cherry (1981), Hardin (1992), and Coolridge-Parker (1989) have shown low correlations between an individual's preference and actual strength. The mode of training selected using a self-report instrument may be among the employee's weakest intake methods.

Research Questions

A number of populations have been tested for modality strengths. The focus of this study is to gather information on the adult basic education population and to compare the modalities exhibited to those of previously tested populations. The research questions to be answered are:

1. What are the modality strengths that appear in adult basic education (ABE) students?
2. How do the modality strengths measured in ABE students compare to other populations?

Methodology

Population and Sample The population selected for this study was the ABE population of Jefferson County Public Schools in Louisville, Kentucky. The sample was taken from the family education centers since the individuals are more consistent in attendance.

Instrumentation and Test Sites The instrument used was the MMPALT-III that consists of seven elements of ten (10) paired stimulus-response items. The MMPALT-III is the latest revision of the original Multi-Modal Paired Associates Learning Test developed by Gilley (1975) and French (1975). The test consisted of six elements: print, aural, visual, interactive, haptic, and kinesthetic. The original version was modified by Cherry (1981) to include an olfactory element and was renamed the MMPALT-II. Cherry and Butler-Tindell updated the test to its current form in 1992. All versions of the test are consistent with Gilley's original recommendations and are consistent with paired associates learning theory (Hardin, 1992). The test is administered to the subjects one element at a time. Both stimulus and response item are given in a predetermined order. The subject is then given the stimulus in a randomized order and is asked to recall the matching response item. The print subtest consists of ten pairs in which a nonsense word is matched with a common word (i.e.,BIW - CAT), the visual subtest consists of ten pairs in which an abstract picture is matched with a common picture, the haptic subtest consists of ten pairs of physical objects, etc. The MMPALT-III determines relative modality strength. Each subtest has a possible range of scores from 0 to 10. Butler-Tindell (1994) determined the scores obtained showed dominant modality capacity (range 8-9-10), moderate modality capacity (5-6-7) and limited modality capacity (0-1-2-3-4). The subjects were tested at their respective instructional locations which were selected at random from ten facilities in Louisville, Kentucky. The test takes approximately one hour to complete.

Reliability and Validity James and Blank (1991) tested 480 adults ranging in age from 19 to 86 and developed the reliability of the MMPALT instrument. The reliability of the subtests was reported to be: print r = 0.85; aural r = 0.80; interactive r = 0.65; visual r = 0.87; haptic r = 0.74; kinesthetic r = 0.67; and olfactory r = 0.73 (Hardin, 1992). Cherry established concurrent validity by statistically comparing the adult population scores with Gilley's original test data (James & Blank, 1991).

Data Collection Informed consent documents were signed before testing. The individuals who participated were from randomly selected locations and volunteered to be tested. Each group was tested following the guidelines set out by the designers of the MMPALT-III. Demographic data was collected at the test site as well. The data collection on the ABE population took place from December 1996 to January 1997 in Jefferson County, Kentucky at the adult education sites. The tests were scored as to the number of correct responses and transferred to a master sheet which contained only the subject's code number, test score, and demographic data.

Data Analysis The data was transferred from each subject's master sheet to SPSS 6.1 for Windows for analysis. The data collected is ordinal in nature and represents a relative strength of modality used by the subject. Mean scores in each modality area were calculated and these means were used to rank order the sample subjects' modalities relative to previously tested populations (Table 1). One sample t-tests were calculated in each modality comparing the ABE sample to previously tested populations (Table 2).

Results and Conclusions

The conclusions that can be drawn from this study are limited at this time by the nature and size of the sample. Data collection will continue to add to the usefulness of the study, however, the basic information to be gained at this time relative to an employee's learning modality is worthy of consideration.

Table 1: Comparison of Rank Order of Modality Strengths

Rank order	Previously tested populations	ABE sample
1	Visual	Visual
2	Interactive	Haptic
3	Haptic	Interactive
4	Aural	Aural
5	Print	Kinesthetic
6	Kinesthetic	Print
7	Olfactory	Olfactory

The rank ordering of modalities indicates that there may be differences in the way individuals receive information. A more useful statistic is the comparison of means between the two groups as seen in table 2.

Table 2: Comparison of Mean Scores

Modal-ity	Test mean	ABE mean	ABE n	SD	SE mean	t-value	df	2-tail alpha
Print	4.71	2.45	20	1.638	.366	-6.17	19	.000
Aural	5.09	2.90	20	1.373	.307	-7.14	19	.000
Interact	5.20	3.10	20	2.614	.584	-3.59	19	.002
Visual	7.84	5.55	20	2.35	.526	-4.36	19	.000
Haptic	5.15	3.90	20	2.693	.602	-2.08	19	.052
Kinesth	3.17	2.50	20	1.701	.380	-1.76	19	.094
Olfactor	1.20	.0.65	20	1.040	.233	-2.37	19	.029

In general it may be stated that ABE employees appear to have less well developed modalities than the comparison population. Early indications are that there is a significant difference in print, aural, visual, interactive, and olfactory modalities at the .05 level. These differences may be mitigated by the low n tested to date.

According to Butler-Tindell (1994), there are three ranges of modality capacity: dominant (range 8-10), moderate (range 5-7), and limited (range 0-4). The ABE population ranges appear to be the same as other populations with the exception of print and kinesthetic. Other groups range from 0 to 10 in print and kinesthetic while the ABE group range is 0 to 7 in print and 0 to 7 in kinesthetic. The range may expand as greater numbers are added to the study.

Contribution to the HRD Field

Employees need to be trained in the most efficient manner possible that the HRD department can provide. ABE level employees may comprise as much as 38% of the work force. If training is to reach the ABE level employee in the most affective manor, it is essential that the professional understand how to approach the instructional task in a way that will be well received. To that end, the MMPALT - III shows early indications of attaining a level of understanding of how an individual receives information. It is then up to the individual HRD department to structure the learning experience to suit the various needs of employees.

One step that may help the trainer save time and money in the training process is testing employees prior to training to determine the way they receive information best. James and Galbraith (1985) observed:

> If adult learners are encouraged to utilize their dominant perceptual learning modality, the learners will be more successful in their learning endeavors whether it be in the classroom or self-directed inquiry (p.20)

If HRD professionals find a tool that accurately assesses the dominant modality of employees, and address this dominance in training, it may be possible to develop more cost effective and more successful programs. The more productive use of money and materials as well as more responsive employees may result.

Studies by a number of researchers have shown learning style modalities to exist in various adult populations (French, 1975), (Cherry, 1981), (Nix, 1983), (Hardin, 1992), (Butler-Tindell, 1994). If working in one's weakest modality can lead to frustration and failure (James, 1991) and working in one's area of strength can enhance success (Grun, 1986), (James,1985) then it would seem important to enable adults to succeed by targeting training to their dominant modality.

References

Butler-Tindell, M.G. (1994). A study of relationships among perceptual modality dominance and choice of college major. Unpublished Doctoral Dissertation, University of Tennessee, Knoxville.

Cherry, C.E. (1981). The measurement of adult learning styles: Perceptual modality. Unpublished Doctoral Dissertation, University of Tennessee, Knoxville.

Coolridge-Parker, J.A. (1989). A comparison of perceived and objectively measured perceptual learning style modality elements of court reporters and court reporting students. Unpublished Doctoral Dissertation, University of South Florida, Tampa.

Cornett,C.E. (1983). What you should know about teaching and learning styles. Bloomington, IN. :Phi Delta Kappa Educational Foundation.

French, R.L. (1975). Teaching strategies and learning. Unpublished Paper, The University of Tennessee, Knoxville.

Gilley, D.V. (1975). Personal learning styles: Exploring the individual's sensory input processes. (Doctoral Dissertation, The University of Tennessee, Knoxville, 1975). Dissertation Abstracts International, 36 (08), 5223A.

Grun, A.F. (1986). An analysis of learning styles and academic achievement based on experiential learning theory, conceptual level theory, and right-left brain hemisphere theory. Unpublished Doctoral Dissertation, University of Minnesota.

Hardin., L.B. (1992). A correlation of learning style and standardized achievement levels of adult vocational students in an area technical center. Unpublished Doctoral Dissertation, University of South Florida, Tampa.

James, W.B., & Galbraeth, M.W. (1985). Perceptual learning styles: Implications and techniques for the practitioner. Lifelong Learning, 8, 20-23

James, W.B., & Blank, W.E. (1991). A comparison of perceptual learning styles of adult high school graduates and non-graduates. Adult Basic Education, 1(2), 98-106.

Jenkins , L.B., & Kersch, I.S. (1994). Adult literacy in Ohio: results of the state adult literacy survey. Columbus, OH: Ohio State Department of Education. Division of Vocational and Career Education. (ERIC Document Reproduction Service No. ED 371 178)

National Institute for Literacy. (1990). What kind of adult literacy policy do we need if we are serious about enabling every adult to become a high skills/high wage worker in the global economy. Washington, D.C.: Policy forum on Achieving the National Educational Goal on Adult Literacy. (ERIC Document Reproduction Service NO. ED 372 278)

Nix, J.H. (1983) The determination of adult learning styles of United States Coast Guard personnel. Unpublished doctoral Dissertation, Florida State University, Tallahassee.

Wislock, R.F. (1993). What are perceptual modalities and how do they contribute to learning? Applying Cognitive Theory to Adult Learning, 59, 5-13.

Measuring Customer Loyalty

Karen A. Brennan
Zenger Miller

This pilot study identified a number of criteria to define loyal and non-loyal customers. They included revenue generated by a customer, satisfaction of the customer, willingness to refer, and time interval since last purchase.

Customer satisfaction can be defined as the measuring of customers' attitudes toward products or services based on the customers' attitude (Stum, 1995). Not only do organizations simply need customers to be satisfied in today's competitive, changing marketplace, they need 'loyal' customers. Obtaining a 'loyal' customer is even more critical than ever before, as the cost of customer defection continues to outweigh the cost of customer retention (Jones and Sasser, 1994). Disloyalty stunts long-term corporate growth and profits by at least 25 to 50 percent (Reichheld, 1996).

Defining and measuring customer loyalty should be an integral part of an organization's "Voice of the Customer" (VOC) system (Farrell, 1996). Other components of a VOC system could include customer satisfaction studies, win/loss studies, and employee satisfaction studies.

In a recent customer loyalty study conducted for a client organization, three main research questions were identified:

1. What are customer buying patterns and trends?
2. How can loyal customers be identified?
3. What do loyal customers want from their suppliers?

The results of this study could be used by the client organization to improve performance by:

- increasing customer retention,
- decreasing customer defection,
- refining and improving marketing, sales, and service strategies,
- providing information for future organization planning, and
- establishing a strong link between customer loyalty survey results and key organizational performance measures.

Methodology

This study was undertaken with an organization in the business services arena. It began with a thorough scan of literature using computer databases available on DIALOG. This resulted in the review of numerous articles, periodicals, and books on the subject of customer loyalty. Next, a number of subject matter experts were interviewed, and national and international conferences were attended to gather background research on customer loyalty issues. Then, existing internal customer-related studies were revisited and compared to various aspects of this study.

Next, the questionnaire, which consisted of 24 items, was developed. Fifty percent of the items were open-ended and 50 percent were closed-ended questions. The instrument was then pilot tested with 25 managers from different organizations. Revisions were made as necessary, and the questionnaire was finalized. The population selected from the organization's customer data base was a random sample stratified by geographic regions throughout North America. To obtain a viable sample of this organization's customer base, a questionnaire was mailed to 1300 customers, and another 500 customers were targeted for phone interviews. Out of these 1800 customers, a total of 322 questionnaires were completed from geographically dispersed areas

© copyright, Karen Brennan, 1997

throughout North America. This resulted in a response rate of nearly 18 percent. Most of the 322 subjects were managers from organizations with 250 or more employees. In addition, three focus groups, consisting of six to eight people per group, were conducted with customers from throughout North America.

Date files and code books were developed which allowed a closer investigation of the data. All available data were entered into the database. Due to the nature of the instrument, both qualitative and statistical quantitative analyses were performed. Categories were developed for the open-ended questions to group 'like' topics. Some of the major categories included Quality of Service, Quality of Products, Meets Customers' Needs, and Corporate Reputation. When statistically significant, a number of questions were cross tabulated with such items as geographic region, industry, size of company, and type of company. Each question was presented with a subjective narrative and corresponding chart. Figure 1 shows a sample chart.

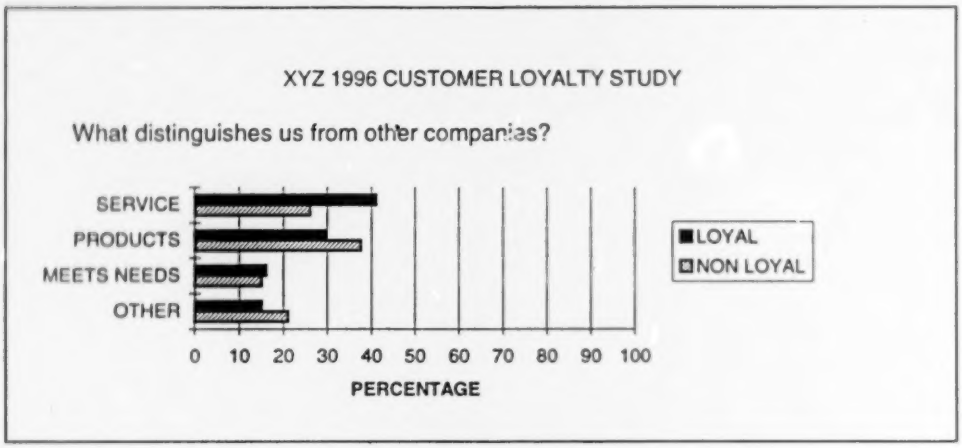


Figure 1. Sample Chart

Findings were identified, and recommendations were provided to key customers within the client organization. Various departments began planning and implementing recommendations that were dependent on the organization's budget and strategy.

Results

Customers were segmented into two groups as loyal and non-loyal customers based on the criteria shown in Figure 2 on the next page:

Loyal Customer Criteria	Non-Loyal Customer Criteria
• among the highest revenue generators	• among the highest revenue generators
• rated their relationship with the organization as a 6 or 7 on a 7-point scale (with 7 being "exceeds my expectations")	• rated their relationship with the organization as a 5 or less on a 7-point scale (with 1 being "does not meet my expectations")
• did feel confident referring the organization inside or outside of their organization	• did *not* feel confident referring the organization inside or outside of their organization
• have purchased products and services from the organization for the last 2 years	• have *not* purchased products and services from the organization for the last 3 years

Figure 2. Loyal and Non-Loyal Criteria

According to the data, this organization has both loyal and non-loyal customers from which a profile for each emerged.

Loyal customers value the organization's quality of service. These customers feel quality of service differentiates this organization from other organizations in the marketplace. Ninety-one percent of loyal customers feel confident referring the organization due to both the quality of its' products and services. In general, these customers are very satisfied with the organization, but urge them to pay attention to the relevance of their products and services to better meet customers' needs. More than half of loyal customers have purchased competitor products and services, and more than half of those customers feel they *have not* had their needs met by these competitors.

Compared to loyal customers, non-loyal customers tend to value high quality products and services of this organization, and they feel that this is what differentiates them in the marketplace. These customers are 15 percent less likely than loyal customers to refer this organization to others. Those who feel confident referring this organization to others tend to feel better about their relationship than those who did not share that confidence. Overall, they are less satisfied with their relationship with the organization. In addition, they see a need for better service from the field service team. They feel the contact frequency from both field and corporate services is insufficient, and they urge the organization to pay more attention to future quality of service. Seventy percent have purchased like products and services from competitors, and they feel these competitors *have* met their needs.

There was no significant difference in customer buying patterns for loyal and non-loyal customers. In general, customers' decisions to *continue* to do business with the organization were for reasons other than quality of service. In most instances, both loyal and non-loyal customers' initial purchase from the organization was for their flagship product. However, these customers also purchase other products and services, but not to a great extent. This observation could mean that more cross-selling opportunities exist for both loyal and non-loyal customers.

Contribution of This Research to New Knowledge in HRD?

Once customer loyalty is defined and measured, organizations can identify competencies needed by employees to ensure customer loyalty and ultimately help contribute to the bottom-line of the organization. For example, in this study, project management skills would enhance the performance of employees both at field offices and at the corporate office. Improved selling skills were also identified as skills that needed sharpening. Improved organizational processes were also identified as an area for organizational improvement.

This pilot study lays the ground work for future customer loyalty studies. A possible later study could compare future revenues to loyal and non-loyal customers. And, finally, this study

could be expanded by applying it to other industries. This would help create a standard instrument.

As competitive pressure, globalization, downsizing, etc. continues to increase, one is sure to find additional press on the topic of customer loyalty on the Internet, in business publications, and in academic books.

Selected References

Farrell, A. (1996), The voice of the customer system, The Farrell Group

Jones, T.O. & Sasser, W.E. (1995, November-December). Why satisfied customers defect. Harvard Business Review, 88-99.

Reichheld, F. (1996). The loyalty effect: The hidden force behind growth, profits, and lasting value, Bain and Company

Stum, D. (1995). Winning customer loyalty in franchise systems. (The Customer Loyalty Monograph Series). Ann Arbor, MI: Customer Loyalty Institute

Textual Analysis with Neural Networks as an Alternative Assessment Tool for Evaluating Responses to Open Ended Questions

William Venable
Oklahoma State University

F. George McDuffee
Oklahoma State University

Discussion of experiences in using neural network based software to evaluate responses to open ended questions on survey instruments. Included are the strengths and limitations of the software package used, problems encountered, "work arounds" for these problems, and suggestions for incorporating this or a similar package for free form text analysis.

One of the more time consuming and irritating aspects of gathering information is using opened ended or unstructured questions. The recording of the data requires that not only the statements of the respondents be accurately transcribed, but when this information is gathered verbally, the body language of the respondent as well as their facial expressions and tone of voice, ethnic background, gender, and age must also be noted if the "true" meaning of their response or reply is to be determined. Even after this information is accurately recorded and transcribed into written format, considerable additional work by highly skilled, trained and experienced evaluators is necessary if meaningful analysis is to be performed. Additionally, if the more quantitative statistical methods are to be used, a coding scheme must be developed and coders trained to transform the raw textual data into numerical data suitable for analysis using standard techniques such as multiple regression.(Raud) Nevertheless, despite these limitations and draw backs, open ended questions, with either written or verbal responses, are frequently the only way to begin an investigation or evaluation where too little is known to permit the creation of closed question or limited response instruments. Indeed, in many cases, one of the goals of the open ended questions is to gain enough background information to allow construction of a useful closed form instrument.

Our original task, as part of a special project with the Oklahoma Department of Corrections to evaluate the effectiveness of vocational education within the Oklahoma penal system, was to evaluate the written explanations, reasons or suggestions to a series of yes/no questions of the form: (Several response lines were provided on the instrument.)

> Did the completion of the program help you stay out of prison? YES NO circle one
> If "YES" how did it help?_____
> If "NO" why didn't it help?_____

Evaluation was complicated by the existence of many overlapping sub-groups of interest within the target/respondent population. For example, male/female, race, age, type of offense, socio-economic background, and native/alien were sub-groups which required sorting and evaluating the responses by these and other categories to determine if any pattern or differences in textual responses could be detected.

The demographic (quantitative) data for the original 948 subjects, supplied by the Oklahoma Department of Corrections in flat file ASCII format was easily imported into DBF format which our statistical package SPSS [Statistical Programs for Social Sciences] could read and it was quickly analyzed. However, as our initial attempts at textual analysis/evaluation quickly showed, because of the volume of material and the large number of comparisons required, traditional manual methods of evaluating textual (qualitative) data were not practicable.

© *William Venable and George McDuffee, 1997*

Main Research Questions

The main question then became "what kinds of computer software, if any, are available to assist in performing qualitative textual analysis," similar to those available for performing statistical analysis on quantitative data. If anything was available, the secondary questions were: (1) what machines/operating systems will it run on; (2) How much does it cost; (3) How easy is it to use; (4) How much computer expertise is required to install and run; and (5) Is the output useful?

Inquiries within our Department, College and University did not indicate knowledge of anything commercially available within a reasonable price range. The WWW was then searched using YAHOO and the other search engines, and the Internet was searched using Gopher, Veronica and Archie. We discovered that the "textual[1] analysis" covers several topics, including comparing one language to another[2] and it is also a technique used to evaluate anonymous manuscripts to determine the author, and to identify additions and modification to old manuscripts such as the books of the Bible. A considerable amount of information was gathered indicating several commercial packages suitable to our particular needs that were available but all of these were very expensive and required powerful computers such as a digital VAX or Sun workstation to run. What appeared to be the suitable packages for our use appeared to be linked with the "buzz words" artificial intelligence, fuzzy logic and neural networks. At about this time, by coincidence, one of our staff attended the International Society for the Scientific Study of Subjectivity[3] [ISSS] conference at Columbia, Missouri. One of the papers presented (Grosswiler), compared and contrasted the results of using Standard Q methodology and a neural network textual analysis program called CatPac to evaluate the classical Q method data set created by Lipsit. (Brown[4] 184-186) While not directly applicable to our problem, the paper did point us in the right direction.

The company manufacturing this textual analysis program was contacted by phone and our problem discussed. It appeared that this software was at least a good start, and a copy was purchased. Later conversation with our colleagues disclosed that several were indeed aware of work that had been done in the area of pattern matching and analysis, but had not connected this to textual analysis. For Engineers and Computer Science people the main interest was in the machine and process control fields. In the social sciences the operative words were "cluster analysis." This underscores the importance of asking the right questions, and not assuming that just because we seem to speak the same language that we necessarily understand each other. This is particularly true when crossing disciplinary/departmental boundaries.

What is Actually Measured?

It should be clearly understood that neural network analysis is not a statistical technique, although some of the same jargon and terms are used. The complete paired comparison similarities matrix (Woelfel 7) is NOT calculated use sum-of-squares, but rather indicates how frequently two words occur "near" each other in the text. The clusters are generated from the common words in associated pairs. For example, if "prjxz" is found to occur frequently in close combination with "ratib," and "ratib" is found to occur frequently in close combination with "zotts," then "ratib," "prjxz," and "zotts" form a common group. The meaning of "close" and "frequently" are adjustable in the programs. CatPac has five optional methods for the identification of clusters. Each method has particular application depending on what types of information the researcher is interested in. (Woelfel 38-39)

---

[1] We also found a significant amount of information under "textural analysis" but this also included the analysis of textures in addition to text.

[2] for additional information see http//www.rut.rice.edu/~barlow/corpus.html

[3] Listserver address Q-METHOD@LISTSERV.KENT.EDU

[4] Email SBROWN@KENTVM.KENT.EDU

What We Discovered

Installation and operation of the CatPac programs was found to be easy and well within the capabilities of anyone with minimal computer skills. It will operate on almost any computer that runs Windows 3.x or Windows 95. (Different programs -- Windows 95 version is better) Huge amounts of RAM and free disk space are not required. However our experience indicates there are other peripheral concerns which should be considered as these have significant effects on how easy to use and how useful the program will be to HRD practitioners and others with large amounts of free form textual data to analyze. Among these concerns are:

The program has no way to select which sections of text within a file to analyze. It simply analyzes all the text in a file. This means that if you wish to compare the responses of Group 1 vs. Group 2, you must create two separate text files. If a large number of categories of interest exist and a large number of comparisons must be made, then it becomes a tedious and error prone process to manually create the necessary files, even if the textual data is not re-input but "cut and pasted" between documents using a word processor. While there may be better solutions, we used an existing database management program called dBaseIV to manage both the quantitative demographic data from the Oklahoma Department of Corrections and the respondents' replies to the survey instrument including their written responses to the open ended questions. With a little programming, text files could be created which contained any desired combination of responses by any desired and identifiable group. For example, the "yes" replies to question 3 by all white males over the age of thirty sentenced to less than 30 months for a non-violent offense.

The program has no way to recognize deviate or alternative spellings, slang/jargon, "non-standard English," nor can it "read between the lines." It is therefore necessary when entering textual data to both correct the spelling and convert into reasonably "standard" English. Additionally, many of the responses were written in "telegraph" format. For example a frequent response was something like "couldn't get no job." Because reflexive usage is important in that it tends to show what the respondent identifies with, and to avoid double negatives, the statement should be reworded "I couldn't get a job." Additionally, indefinite pronouns such as "it" and "that" are very context sensitive, and can mean very different things even in two adjacent sentences by the same respondent to the same question. Therefore, for proper analysis, it appears to be necessary for the person doing the text transcription (input into the computer) to "translate" indefinite pronouns into defined and consistent nouns. Otherwise there is the risk that "it" will be a major concept/word that is actually referring to several different things, or where different synonyms are used for the same thing, important words and relationships/clusters may not be discovered. While this seems like an easy task to do while entering the data, in practice it was found to be quite error prone and time consuming. Additionally, the person doing the "translation" of slang/jargon must understand what is meant by the expressions and idioms used.[5]

The program seems to handle simple declarative sentences best. It is unknown how ironic or sarcastic responses should be identified or handled. Plural usage may also be a problem. Frequently in our material, the singular terms inmate, graduate or prisoner were self reflexive while the plural terms inmates, graduates, or prisoners tended to refer to everyone else. Where this is not the case, the program provides a global "find and replace" option to "consolidate" usage.

While the program outputs are detailed and well organized, their meaning is not particularly clear, at least to us. It is not that there is any particular difficulty in seeing a

[5] Such "editing" and "transliteration" caused about a 1 grade increase in average reading level in the text material, from roughly 6th to 7th grade.

difference in responses between groups, for example the words "drugs" and "alcohol" showed up much more frequently in responses from the recidivist groups than the non-recidivist groups, the problem is that the significance of these differences is unknown. In many ways this corresponds to having a statistical program that will generate the mean and standard deviation from groups of quantitative data, but not having the knowledge to perform a *t* test to determine if the differences are significant. Indeed, it does not appear to be known if such significance evaluation or testing is even theoretically possible. It may well be that the best that can be done in "natural language" analysis is to provide some type of descriptive statistics where a large amount of textual data is consolidated into a few key words and how these cluster.

The Oresme interactive non-hierarchical cluster analysis feature seems to provide considerable insight into how the respondents in a group tend to associate concepts (words). In practice this is two copies of the most common words[6] in the text file. When you select a word in the first list, words that are associated with that word are highlighted in the second list. A numerical value indicating the amount of association is also indicated. You can select more than one word in the first list, for example "drugs" and "alcohol" to see which words in the second list are strongly associated with either of those words. The words in both lists are the ones selected by the CatPac program. These can provide direction and suggestions for analysis. For example, while recidivists and non-recidivists may tend to use the about same words to respond to particular questions as determined by the word frequency analysis, Oresme indicates that the words tend to cluster quite differently. The non-recidivists may tend to associate "I," "job," "opportunity" and "family," while the recidivists may tend to associate "I," "job," "can't," and "problem."

Sample Input Files and Analysis Outputs (Spelling and grammar corrected as indicated above)

Inmate responses to "Did completion of the program help you get a job and if 'yes' how did it help?"

I had experience because of the VOTECH program. The fact that I was a VOTECH graduate. Because I had good comprehensive VOTECH training. Because without the VOTECH program I would not have had the skills and knowledge. I knew a lot more about painting and carpentry work because of the VOTECH program. I had experience because of the VOTECH program. I had the experience required on the job because of the VOTECH program. If I hadn't taken the VOTECH program I wouldn't have known how to do the job. Having a VOTECH certificate and basic knowledge of equipment and tools was main reason for employment. First the VOTECH program provided me with the skill in which I trained for and I was informed of the advantages I could gain for my future. I knew how and what to do on the job because of the VOTECH program. Because of the VOTECH program I knew the tools and everything. I had on the job experience through the VOTECH program. The VOTECH program gave me knowledge of measurements, weights and metal densities. The VOTECH program taught me how to follow instructions. The VOTECH program taught me how to direct my self with respect. The VOTECH program gave me the proper skills for a book keeping job. The VOTECH program gave me typing & computer skills. After completing the VOTECH course in automatic transmission I had a lot more self esteem but I must admit that the course I took in "Breaking Barriers" helped me very much. It was an experimental course that came with the class I took. The first place that I stopped thought it very good that I had received a certificate of excellence from the VOTECH class. I was able to put down more pertinent information than N/A. They were specifically hiring heat and air trained people. I figure from gaining the knowledge in the various areas I did in the

[6] CatPac has an exclude facility to exclude common words such as the, and, but, etc. and this list can be customized for your particular documents. Also specific words of interest can be "forced" into the analysis even if these are infrequently used.

VOTECH class it helped very much towards getting this job. The VOTECH program provided me with basic experience. By having a higher VOTECH education I showed I was willing to learn new areas. The information and knowledge in the landscaping VOTECH helped me know the seasons, names of plants and products. The VOTECH program helped me by preparing me with skills needed to work as a mechanic. The VOTECH program helped me to be more persistent in getting a job. The VOTECH program was what the job called for. The VOTECH program gave me knowledge about plumbing and electrical, but I need more electrical training. The VOTECH program prepared me with the knowledge I needed to do the job. They liked it because I went to VOTECH. VOTECH prepared me with the knowledge and hands on experience to begin a career in the field of auto mechanics. The VOTECH program built my self esteem and I was prepared for any job they had to offer. The VOTECH program prepared me to be a valuable member of society and gave me self esteem. The VOTECH program gave me hands on experience of working with motors and stuff. The VOTECH program provided me with a working knowledge related to the field. The VOTECH program gave me some on the job training and book knowledge. The VOTECH program gave me more confidence in myself. I went to work for my VOTECH instructor's son. I started out doing general maintenance and ended up cooking and management work.

CatPac Analysis of above text

TOTAL WORDS	159	THRESHOLD	0.000
TOTAL UNIQUE WORDS	10	RESTORING FORCE	0.100
TOTAL EPISODES	153	CYCLES	3
TOTAL LINES	58	FUNCTION	Sigmoid (-1 - +1)
		CLAMPING	Yes

DESCENDING FREQUENCY LIST

WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
VOTECH	38	23.9	148	96.7
I	32	20.1	98	64.1
PROGRAM	27	17.0	127	83.0
ME	21	13.2	101	66.0
JOB	11	6.9	73	47.7
KNOWLEDGE	10	6.3	67	43.8
EXPERIENCE	7	4.4	41	26.8
HELPED	5	3.1	30	19.6
PREPARED	4	2.5	25	16.3
SELF	4	2.5	28	18.3

ALPHABETICALLY SORTED LIST

WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
EXPERIENCE	7	4.4	41	26.8
HELPED	5	3.1	30	19.6
I	32	20.1	98	64.1
JOB	11	6.9	73	47.7
KNOWLEDGE	10	6.3	67	43.8
ME	21	13.2	101	66.0
PREPARED	4	2.5	25	16.3
PROGRAM	27	17.0	127	83.0
SELF	4	2.5	28	18.3
VOTECH	38	23.9	148	96.7

(Dendogram moved to end of this section for ease in comparison)

Inmate responses to the question "Did completion of the program help you get a job and if 'no' why didn't it help?"

Job wasn't related to horticulture type work. Job was not in the line of skill I was trained for. Job wasn't related to skill I trained for. I really didn't learn nothing. The program prepared me very well to do a good job, but DOC released me with $190 and barely a place to live. Also most employers didn't want an ex-con and it's on my record. I applied for over 20 jobs. The VOTECH program just didn't help. Job that I applied for would not accept me without years of experience and my prior convictions were a problem. I never applied for a plumbing position. I feel that I didn't have the proper contact that I needed so the training was wasted. Job I got was not related to VOTECH courses taken. Because people in general have a complex about hiring convicted felons and it's hard for anyone who has a felony to get a job. Job wasn't same trade I trained for. I learned something I like doing. I have done landscaping & horticulture work at parks, small towns, and Cowboy Hall of Fame. Because my family owns a carpet store and that's where I went to work. I took welding not cooking in VOTECH but that's what I am doing. Education isn't necessary at a saw mill, but I worked at the sawmill to build a job history. I combined that job

and going to VOTECH to get other jobs. Completing VOTECH didn't help me on my first job but VOTECH did on other jobs. I had this job previous to jail time. I re-opened my old company. The job that I got was at a restaurant. I already had this job setup. VOTECH program was not in the same field as the job. I didn't get a job because I got heart trouble but I think the program would have helped. There are many questions on why didn't someone of many programs available for inmates being released help me. But as you know or should know not one person after so many inquiries would help me. My lifelong trade was unrelated to the VOTECH wood working program, however should a need arise, I have this new avenue open to me now. It's hard for a person who is new at a certain skill to get a full grasp on all that goes with that skill in the amount of time that the Oklahoma DOC gives an inmate to complete their VOTECH program -- it's rush, rush, rush. At that time I didn't have any transportation and no jobs were available that used my VOTECH skills. VOTECH program didn't give enough information on companies who would hire semi-qualified personnel. Job was not related to my program -- I needed further training. The job did not require the use of diesel mechanics. Had to take job although the job was not what I was trained for.

CatPac Analysis of above text

TOTAL WORDS	96	THRESHOLD	0.000
TOTAL UNIQUE WORDS	10	RESTORING FORCE	0.100
TOTAL EPISODES	90	CYCLES	3
TOTAL LINES	47	FUNCTION	Sigmoid (-1 - +1)
		CLAMPING	Yes

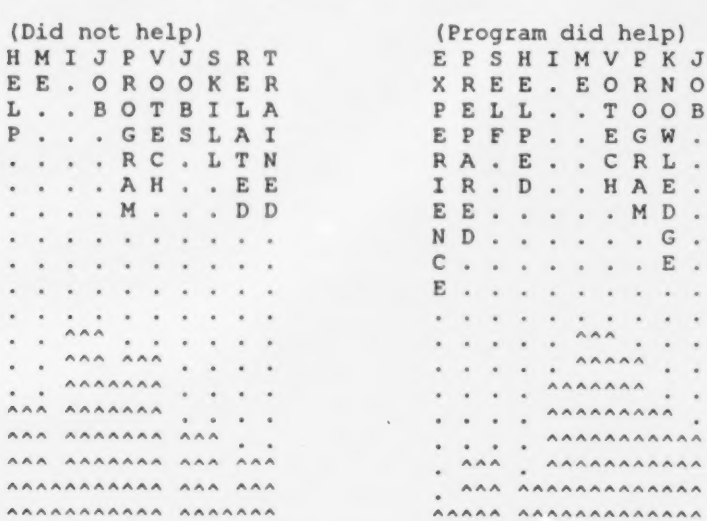
DESCENDING FREQUENCY LIST

WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
I	30	31.2	85	94.4
JOB	20	20.8	71	78.9
VOTECH	11	11.5	58	64.4
PROGRAM	8	8.3	47	52.2
ME	7	7.3	36	40.0
HELP	4	4.2	23	25.6
JOBS	4	4.2	27	30.0
RELATED	4	4.2	22	24.4
SKILL	4	4.2	17	18.9
TRAINED	4	4.2	19	21.1

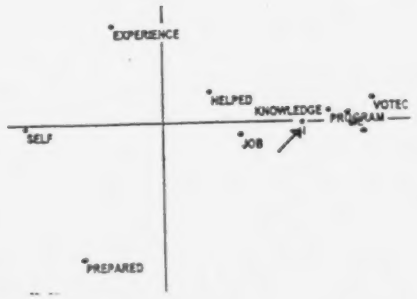
ALPHABETICALLY SORTED LIST

WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
HELP	4	4.2	23	25.6
I	30	31.2	85	94.4
JOB	20	20.8	71	78.9
JOBS	4	4.2	27	30.0
ME	7	7.3	36	40.0
PROGRAM	8	8.3	47	52.2
RELATED	4	4.2	22	24.4
SKILL	4	4.2	17	18.9
TRAINED	4	4.2	19	21.1
VOTECH	11	11.5	58	64.4

WARDS METHOD Dendograms

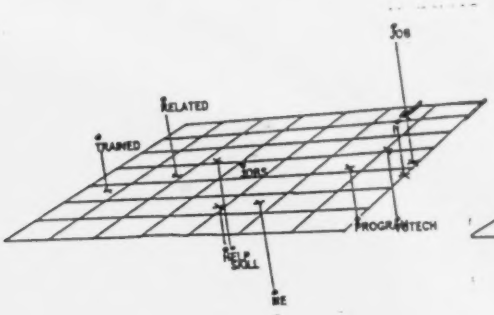

(Did not help)
H M I J P V J S R T
E E . O R O O K E R
L . . B O T B I L A
P . . . G E S L A I
. . . . R C . L T N
. . . . A H . . E E
. . . . M . . . D D
(Program did help)
E P S H I M V P K J
X R E E . E O R N O
P E L L . . T O O B
E P F P . . E G W .
R A . E . . C R L .
I R . D . . H A E .
E E . . . . . M D .
N D . . . . . G .
C . . . . . E .
E . . . . . .

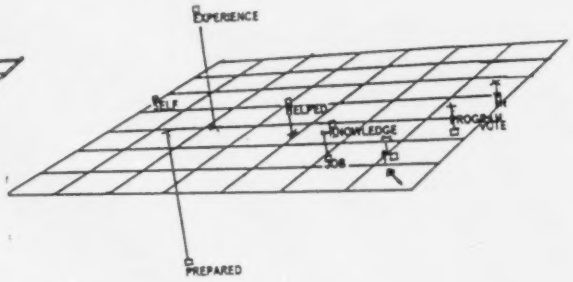
Two dimensional graphical representations of above word relationships rotated so that "I" is on the positive X axis for ease of comparison.



(Did not help) **(Program did help)**

Three dimensional graphical representations of above word relationships rotated so that "I" is at the right hand side of the graph:





(Did not help) **(Program did help)**

Conclusions

CatPac can not replace existing methods of textual analysis in all situations, however it has the potential to greatly assist the researcher by consolidating masses of material into smaller pieces that are more easily understood.

The Oresme interactive non-hierarchical cluster analysis feature is very useful in indicating possible lines of investigation and study. While Oresme will indicate what words are associated within a cluster, it does not explain how or why, therefore the investigator must go back to the original text document and locate the words for "manual" evaluation. The CatPac "find" function is very helpful for this.

If the intent of the researcher is to determine if and to a limited degree how known or posited identifiable groups differ in free text responses to open ended questions than CatPac appears to be a useful tool. However if the intent is to determine if sub-groups with unknown membership criteria exist with different perceptions and attitudes, particularly if the ultimate goal is to identify the salient beliefs of these sub-groups, then other methodologies such as Q will most likely be a better choice.

CatPac does not function as a psychoanalyst, therefore totally free stream-of-conscience material is difficult to analyze and compare. Simple declarative statements in (pertinent) response to specific but open ended questions seem to work the best.

Recommendations

If you or your organization has the need to qualitatively analyze large amounts of textual material and/or perform a large number of comparisons among/between different groups then, subject to the above limitations, CatPac or similar neural network software may well be of considerable help to you.

As with any new software or methodology, it should not be initially used on a critical or time sensitive project. Several hours of experimentation with the program spread over a few weeks are highly suggested. We have found a good way to develop a feel for what this software can and cannot do is to analyze your writing, such as a report or memo that you have just completed while the material is still fresh in your mind. Another approach is to re-analyze textual material and compare the results with the first analysis.

Bibliography

Brown, S. R. (1980) Political Subjectivity: Applications of Q Methodology in Political Science. Westford, Mass: Murray Printing. ISBN 0-300-02579-3

Grosswiler, P. (1996, October 3-5) Neural network analysis as an alternative to factor analysis in the mapping of intersubjectivity in Q-methodology communication research. Paper presented at the Twelfth International Conference on Subjectivity. International Society for the Scientific Study of Subjectivity, School of Journalism and Stephenson Research Center, University of Missouri-Columbia.

Raud, R. and Falling, M. (downloaded 10/07/96) Automating the Coding Process with Neural Networks. http://www,monmoth.com/~rraud/autocode.html

Woelfel J. K., et al. (1995) CatPac for Windows User's Manual Michigan: Terra Research & Computing Company. Information also available at http//www.terraresearch.com/

The Impact of Influential Observations on Multiple Regression Analysis in HRD Research

Reid A. Bates
Elwood F. Holton III
Michael F. Burnett
Louisiana State University

An overlooked aspect of multiple regression research using employee provided data is the potential for micro level bias that can unduly influence regression results. This article demonstrates the potential impact of influential observations using a case study. In this case, the identification and deletion of only three observations changed the regression model from non-significant to robust with significant theoretical implications. Procedures for identifying influential observations are outlined.

Multiple regression analysis is a widely used analytical technique in HRD research, particularly in studies attempting to predict outcomes of HRD interventions. Recent examples of studies employing multiple regression include those focusing on transfer of training (Gielen, 1995; Rouiller & Goldstein, 1993; Xiao, 1996), new employee development (Holton & Russell, 1996), factors affecting the opportunity to perform trained task on the job (Ford, Quinones, Sego, & Sorra, 1992), organizational feedback environment and performance (Becker & Klimoski, 1989), age, work attitudes, and performance (Cleveland & Shore, 1992), trainee characteristics and outcomes of open learning (Warr & Bunce, 1995), training fulfillment, organizational commitment, self-efficacy, and motivation (Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991), and the influence of pretraining experiences on learning in training (Smith-Jentsch, Jentsch, Payne, & Salas, 1996).

Field studies such as these rely on data collected from employees in organizations. An easily overlooked aspect of using data collected in the field is the tremendous potential for biased responses which can unduly influence regression results. Although large-scale bias (i.e., that in which a large number of responses may, for some reason, be invalid) is a prominent concern for most researchers, bias which affects only a relatively small number of responses is often left unexplored. This article demonstrates that small-scale bias may be a significant issue when using regression analysis in HRD research.

The goal of regression analysis is to estimate the most representative statistical model given certain sample data. This model is presumed to be the one that best reflects the population from which it was drawn and provides the most valid and generalizable results (Darlington, 1990). The presence of one or several influential observations in a sample can undermine achievement of this goal. Influential observations are observations with a disproportionate influence on one or more aspects of the estimates in a regression analysis. This influence may arise as a result of (a) substantial differences in the predictor values from other cases, (b) extreme values for the criterion variable(s), or (c) a combination of these effects (Hair, Anderson, Tatham, & Black, 1995).

The impact of influential observations can be harmless in the sense that they reinforce the pattern of data and therefore do not unduly distort regression results. Alternatively, when influential observations introduce bias into the regression results their effect can be pernicious. In the latter case, deletion of such observations may be desirable in terms of the statistical properties of the parameter estimates as well as the estimate of the final equation (Hocking, 1976). The function of a diagnostic analysis is to identify influential observations and to determine whether those observations should be excluded from the final regression analysis.

This purpose of this paper is to (a) illustrate, using a sample data set, how a small number of observations can significantly influence the results of a regression analysis; and (b) describe a straightforward diagnostic analysis methodology for identifying influential observations in regression analysis, determining the relative influence of these observations on regression estimates, and

© *copyright 1997, Bates, Holton & Burnett*

selecting which observations to delete from data analysis.

Method

Sample The subjects for the present study were production operators in two production units of a large petrochemical manufacturer. Production operators in these units were responsible for monitoring, operating, and maintaining production equipment used to produce a variety of chemical products. Operators were required to complete from 30 to over 100 computer-based training modules depending on their job responsibilities in order to meet organizational requirements as well as federal certification mandates for the management and handling of highly hazardous chemicals. Following the completion of training in May, 1995, the data collection instruments used in this study were administered.

Measures The independent variables examined in this study were learning, organizational commitment, performance utility, perceived training content validity, transfer design, and transfer climate variables including positive personal outcomes, negative personal outcomes, peer support, supervisor sanctions, supervisor support, opportunity to use, and change resistance. A battery of instruments were developed or adopted to collect self-reports of these and other factors believed to affect the transfer of training process. For all instruments, common factor analysis with oblique rotation was used to identify latent constructs and scales.

The dependent variable used in this study was supervisor ratings of job performance. The performance measurement process involved identification of a subset of critical tasks on which operator performance was frequently observed by supervisors, and collection of supervisory judgements of percentage of time these tasks were done correctly. In addition to the development of elaborate procedures to collect criterion data, steps were also taken to insure content validity of the performance rating instruments. Due to space limitations a complete description of these procedures is not presented here (see Bates, 1997).

Analysis The relationships between independent and dependent variables were analyzed using bivariate correlation analysis and hierarchical regression analysis. A more complete discussion of the procedures and results can be found in Bates, Holton & Seyler (1997).

Regression Diagnostic Procedures. Procedures for regression diagnostics can be found in most regression texts. We found the procedures in Hair et al. (1995) to be most useful and adopted them as a guide in our analysis. The diagnostic process employed in this study was conducted in these steps:

1. Run initial regression analysis.
2. Test for violation of regression assumptions.
3. Test for multicollinearity.
4. Test for individual influential observations.
5. Test for effect of multiple influential observations.

Run Initial Regression Analysis. Since a diagnostic analysis always addresses a particular regression, the first step in diagnostic analysis is to identify the regression to be run, assuming no irregularities (Darlington, 1990). The diagnostic analysis in this study focused on the full regression model which included twelve predictors (organizational commitment, training content validity, performance utility, learning, transfer design, positive personal outcomes, negative personal outcomes, peer support, supervisor sanctions, supervisor support, opportunity to use, and change resistance), with performance ratings as the criterion and an initial regression sample size of 65.

Test for Violation of Regression Assumptions. In conducting multiple regression analysis, several assumptions are made about the relationships between criterion and predictor variables that can affect the least squares statistical procedure. The four assumptions to be addressed include (a) linearity of the relationship between criterion and predictor variable(s), (b) constant variance of the error terms (homoscedasticity), (c) normality of the error term distribution, and (d) the independence of residuals (Hair et al., 1995). The extent to which the data confirmed these assumptions was evaluated by examining various residual plots. Satisfying these assumptions helps to ensure that the research findings are representative of the sample and that the best results possible have been obtained (Freund & Wilson, 1993).

Multicollinearity A key issue that can have a substantial impact on the results of multiple regression analysis is the degree to which predictor variables are correlated with each other, i.e., the degree of multicollinearity among predictors. Multicollinearity can (a) limit the size of the coefficient of determination (R^2) and make it more difficult to increase unique explanatory prediction from additional variables, and (b) make determination of the unique contribution of each predictor difficult to assess because the effects of the predictors are confounded due to their intercorrelation. In short, a high degree of multicollinearity means larger portions of shared variance and lower levels of unique variance (Hair et al., 1995).

Assessment of multicollinearity in the present study followed the methodology developed by Belsley, Kuh, and Welsch (1980). This process utilized (a) the condition index, an index that represents the collinearity of combinations of predictor variables; and (b) the regression coefficient variance-decomposition matrix which describes the proportion of variance for each regression coefficient attributable to each condition index. All of these measures were available as standard options in SPSS and SAS. The multicollinearity assessment process involved identification of all condition indices above a threshold value (30) and identification of variables within those condition indices that account for a substantial proportion (.90 or greater) of the variance for two or more coefficients. Variables with a condition index greater than 30 and which account for .90 or greater of the variance of two or more coefficients are considered to exhibit multicollinearity. This process is reported to provide greater diagnostic power in the assessment of multicollinearity than the more typical methods examining tolerance values or variance inflation factor (VIF) values (Hair et al., 1995).

Testing for Individual Influential Observations. Diagnostic analysis in regression aims at identifying influential observations and involves efforts to detect three classes of influential observations. These include (a) outliers on the dependent variable; (b) leverage points, or outliers on one or more independent variables; and (c) single observations capable of influencing the regression results in terms of the residuals, individual coefficients, the coefficient matrix, or total model fit (Hair et al., 1995). Common methods used to detect these different classes of influential observations include examination of studentized residuals to identify outliers, centered leverage values or Mahalanobis distance for leverage points, and single case diagnostics such as DFBETA values, studentized deleted residuals, and Cook's distance. Again, these measures are available as standard options in SPSS and SAS.

Test for Influence of Multiple Influential Observations. All of the techniques discussed above are capable of identifying individual influential observations and their estimated impact on regression results. They do not, however, provide information about the potential influence on the regression analysis of two or more influential observations in combination. In order to assess the relative impact on the regression analysis of multiple influential observations in combination, a variable selection technique using all-subsets regression is useful (see Hocking, 1976; Peixoto & LaMotte, 1989). This analysis identifies the maximum-R^2 subset regression model of each size (e.g., subset regression models with one influential observation deleted, two, three and so on to include all individual influential observations). Subsequent to the identification of a maximum-R^2 subset regression model for each subset size, comparative evaluation of the maximum-R^2 subsets follows a technique described by Hocking (1976). This techniques calculates an F-statistic using the difference in residual sum of squares for each model to identify which subset(s) offer significant improvements in model prediction and estimation. This F-statistic is computed using the formula

$$F_d = \frac{\text{Residual SOS}_A - \text{Residual SOS}_B}{\sigma^2}$$

where Residual SOS_B is the residual sum of squares of the model with X observations deleted, Residual SOS_A is the residual sum of squares of the model with X - 1 observations deleted, σ^2 is the residual mean square for the full model (no deletions), and df = 1, n -1 - total predictors - total outliers being tested (Peixoto & LaMotte, 1989). This is the only statistic that had to be calculated

by hand, but all of the variables in the formula were available on the standard regression printouts.

This approach postulates that the last model in a progressive comparison procedure to produce a significant F-statistic indicates the appropriate combination of influential observations to be deleted from the analysis (Hocking, 1976). Evaluating the impact of multiple influential observations in this manner provides a more discretionary approach to the identification and deletion of unrepresentative observations than that based solely on identification of individual influential observations.

Results

Step 1: Multiple Regression Results With Full Data Set The results of a six step hierarchical regression run with the full data set showed (a) none of the individual steps in the regression produced a significant model ($p \leq .05$); (b) none of the increases in explained variance produced by the addition of variables to the model were significant, including the addition of the transfer climate variables as a set; (c) the full model with all predictors entered was not significant ($F = .97$ ns), and did not explain a significant proportion of the variance in performance ratings ($R^2 = .18$ ns); and (d) only one variable, peer support, emerged with a significant Beta value ($\beta = .41$, $p \leq .05$). These results are summarized in Table 1.

Table 1: Results of the Multiple Regression With the Full Data Set

Step	Variable Entered	R^2 Adj R^2	F_{model} (df)	ΔR^2	F_{error} (df)	β''
1	Organizational Commitment	.004 -.012	.268 (1,63)	—	—	-.055
2'	Content Validity	.014 -.018	.424 (2,62)	.01	.529 (1,62)	.06
3'	Performance Utility	.027 -.021	.569 (3,61)	.013	.818 (1,61)	-.108
4'	Learning Average	.031 -.033	.486 (4,60)	.004	.248 (1,60)	-.04
5'	Transfer Design	.043 -.038	.533 (5,59)	.012	.741 (1,59)	-.096
6'	Supervisor Sanctions Pos Pers Outcomes Neg Pers Outcomes Change Resistance Peer Support Supervisor Support Opportunity to Use	.183 -.006	.97 (12,52)	.14	1.27 (7,52)	.315 -.029 .034 .118 .409* .009 .035

' Each successive model in the hierarchical regression analysis includes the variables previously entered.

'' Beta values shown were calculated with all predictors entered in the regression equation.

* $p \leq .05$

Step 2: Test of Regression Assumptions For the purposes of this paper, these analyses will not be discussed in full due to space limitations. They indicated, however, that no serious violations of the basic assumptions of multiple regression analysis occurred.

Step 3: Multicollinearity Assessment. Analysis of the data indicated no substantial multicollinearity problem among the predictor variables.

Step 4: Test for Individual Influential Observations Identifying individual influential observations involved examination of several diagnostic statistics:

1) Outliers on the dependent variable were identified through examination of studentized residuals. Studentized residuals with values greater than ± 1.96 are significant. As shown in Table 2, five observations were identified (numbers 41, 45, 46, 53, 68) as outliers. Identification of outliers is important because outliers typically do not fit the regression equation well, may have an influential effect on the regression results, and may therefore require some remedy (Hair et al., 1995).

BEST COPY AVAILABLE

2) Leverage points in the data (i.e., an observation that substantially influences regression results based on its difference on one or more independent variables) were identified using centered leverage values. Centered leverage values greater than *2(# of predictors) + n* for sample sizes greater than 50 may be influential observations (Hair et al., 1995). For the present data set, this threshold value was .369. Four leverage points (34, 49, 54, 68) were identified (see Table 2).

3) Single case diagnostics delete an individual case from the regression estimation and provide a measure of the change in results. Several were examined in this study. Examination of studentized deleted residuals yielded significant values (> ± 1.96) for five observations (41, 45, 46, 53, 68). These are the same observations identified in step 1, the residual analysis. In this instance, however, these observations are outliers from the regression equation which has been estimated *without* their influence (Hair et al., 1995). No exact threshold values are available for DFBETA values. Therefore, examination of a plot of DFBETA values is helpful in identifying observations having substantial impacts on the coefficients. As shown in Table 2, examination of the DFBETA plot showed two observations (46, 68) to have a substantially greater influence on the intercept than other observations. Cook's distance captures both the leverage effects and the change in residuals associated with single observations. That is, it provides a summary measure of the influence of a single observation based on the changes in all other residuals when the case is omitted from the regression estimation. Threshold values greater than one are generally regarded as indicative of substantial influence for a case on the estimated regression coefficients. However, "if no observations exceed this threshold, additional attention is dictated if a small set of observations has substantially higher values that all of the remaining observations" (Hair et al., 1995, p. 157). In this data, no single case had a Cook's value greater than 1. However, three cases (41, 46, 68) had values which substantially exceeded the overall mean (.003). These cases were therefore judged to be deserving of further analysis.

Table 2: Results of Diagnostic Analysis for Individual Observations

Observation	Studentized Residual	Centered Leverage Value	Studentized Deleted Residual	DFBETA	Cook's Distance
34		.41			
41	-2.72		-2.91		.22
45	-2.87		-3.09		
46	-2.44		-2.57	-1.66	.24
49		.43			
53	-2.64		-2.81		
54		.41			
68	2.99	.46	3.26	1.08	.62

Based on the results of the diagnostic analysis, five influential observations (41, 45, 46, 53, & 68) were selected for further analysis. These observations were identified as falling outside the general pattern of data and were identified as influential observations by two or more diagnostic measures. This suggests these observations have the greatest potential influence in terms of improving our ability to estimate a regression equation on a representative sample and to obtain the most externally valid results. Observations 34, 49, or 54 were not selected because, although these cases were identified as having an influential impact on the regression results as indicated by their leverage values, they fall in the general pattern of the data and therefore do not threaten to distort the regression results.

Step 5: Test for Effect of Multiple Influential Observations The final step in the diagnostic analysis involved assessing the relative impact of these five observations individually and in combination on model estimation. This was done using an all subsets regression (see Hocking, 1976; Peixoto & LaMotte, 1989). With five influential observations being examined, a total of 32

possible regression models were identified with various subsets of zero, one, two, three, four, and five variables deleted. From each subset of models with deleted variables, the model with the maximum R^2 was identified. To compare the relative effect of these maximum R^2 subset models on regression estimation, an F-statistic was computed for the difference in residual sum of squares for the five models plus the full model (no deletions).

This approach assumes that the last model in a progressive comparison procedure to produce a significant F-statistic indicates the appropriate combination of influential observations to be deleted from the analysis (Hocking, 1976). As seen in Table 7, the last significant ($p \leq .05$) F-statistic occurs when the model with two observations deleted is compared with the model with three observations deleted. This suggests that the model with three observations deleted (observations 41, 45, & 68) is the most appropriate model for the final regression analysis. Deletion of these observations reduced the final sample size to 62.

Table 3: Comparison of F-statistics for Maximum R^2 Subsets

Models Compared	$F_{1,41}$
Full Model v Delete Case 41	10.32*
Delete Case 41 v Delete Cases 41 & 68	11.72*
Delete Cases 41 & 68 v Delete Cases 41, 45, 68	4.99*
Delete Cases 41, 45, 68 v Delete Cases 41, 45, 53, 68	3.91
Delete Cases 41, 45, 53, 68 v Delete Cases 41, 45, 46, 53, 68	3.67

* $p \leq .05$

Step 6: Multiple Regression Results With Influential Observations Deleted Several significant changes in model estimation are apparent with the deletion of three influential observations (41, 45, 68). First, although none of the steps in the initial regression analysis with the full data set produced a significant increase in explained variance, two steps in the modified analysis are worth noting. The introduction of content validity at the second step in the regression produced a 5.9% increase in explained variance, a finding that approached significance ($F = 3.73$, $p \leq .06$). The data also showed that, when entered as a set, the transfer climate variables increased the total variance explained in the model by . 36, a significant gain ($p \leq .001$). Second, in contrast to the initial analysis, with all of the predictors entered, the modified analysis produced a significant regression model ($F_{12,49} = 3.47$, $p \leq .001$) which explained a substantial proportion ($R^2 = .46$) of the variance in performance ratings. Finally, examination of Beta values indicated four significant coefficients as compared to one in the initial analysis. As seen in Table 4, with all variables entered in the regression model, content validity ($\beta = .44$, $p \leq .01$), supervisor sanctions ($\beta = .38$, $p \leq .01$), change resistance ($\beta = .38$, $p \leq .01$), as well as peer support ($\beta = .52$, $p \leq .001$) emerged with significant Beta values.

Discussion

This case study demonstrates the potential dangers associated with organizational research where biased data may occur. In this case, the conclusions reached from the initial analysis would have been dramatically different if the diagnostic procedures had not been implemented. The conclusion from the initial regression model would have been that no transfer of learning occurred, no significant regression model was estimated from the predictors, and there was only one significant predictor. Instead, with the deletion of a relatively small number (3) of influential cases, a robust regression equation emerged in which 46% of the variance in transfer was explained by a subset of predictors. Moreover, these findings (a) highlighted the value of several unexplored dimensions of transfer climate, (b) spawned important implications about the role of transfer climate variables in training effectiveness, and (c) suggested an important modification of the theoretical model used in the study.

Table 4: Results of the Multiple Regression With Influential Observations Deleted

Step	Variable(s) Entered	R^2 Adj R^2	F_{model} (df)	ΔR^2	F_{change} (df)	β''
1	Organizational Commitment	.004 -.013	.233 (1,60)	—	—	-.181
2'	Content Validity	.063 .032	1.99 (2,59)	.059	3.73 (1,62)	.443*
3'	Performance Utility	.07 .022	1.45 (3,58)	.007	.437 (1,58)	.034
4'	Learning Average	.082 .017	1.27 (4,57)	.012	.75 (1,57)	-.117
5'	Transfer Design	.10 .02	1.25 (5,56)	.018	1.13 (1,56)	-.055
6'	Supervisor Sanctions	.46	3.47**	.36	4.64**	.387*
	Pos Pers Outcomes	.327	(12,49)		(7,49)	.095
	Neg Pers Outcomes					.097
	Change Resistance					.382*
	Peer Support					.521*
	Supervisor Support					-.032
	Opportunity to Use					-.211

' Each successive model in the hierarchical regression analysis includes the variables previously entered.

'' Beta values shown were calculated with all predictors entered in the regression equation.

* $p \leq .01$ ** $p \leq .001$

It is important to note that use of all subsets regression to assess the impact of combinations of influential observations on regression results led to the elimination of three observations rather than five as indicated by analysis of the individual impact of potentially influential observations. The results of the present research suggest that incorporation of this methodology into diagnostic analysis may be valuable insofar as it provides HRD researchers with greater power to be discriminating when making decisions about case deletion in regression analysis.

It would be tempting to conclude the analytical situation portrayed here is an unusual case when, in fact, it is probably quite typical. HRD field research often asks employees to provide data they could perceive as potentially damaging to their own or someone else's career. Typical examples include opinions about a supervisor, opinions about management, perceptions of peers, and supervisory perceptions of subordinate behavior. Although good organizational survey development procedures include steps to prevent large scale mistrust and socially desirable responses, this study demonstrates that this macro level bias is only part of the potential problem. As seen in this study, micro level bias may occur when even a very few employees respond in an odd manner. Furthermore, micro level bias can dramatically alter the results obtained, particularly with small sample sizes as often occur in HRD research.

The very nature of HRD research in organizations suggests that most studies are at risk for some level of suspect responses due to a variety of naturally occurring circumstances. One possible reason is that there is almost always some level of concern, perhaps mistrust, of organizational surveys, especially when the data may be potentially damaging to an individual employee. Despite all best procedures and efforts by researchers to reassure employees of confidentiality, it is likely that some employees may choose to protect themselves. It is probably impossible to prevent a few employees from being reluctant to provide valid answers to the instrument.

Another likely possibility is that interpersonal affect may influence employee responses. In discussing the role of affect in performance appraisal, Cardy and Dobbins (1994) suggest, for example, that differentiated liking (i.e., liking directed toward a particular ratee) may impact perceptions of performance both (a) directly insofar as raters are not able to separate affect from actual performance, and (b) indirectly through its influence on schema formation or attribution formation. Still another possibility is the occurrence of some organizational event which affects

only a few employees in such a way that it changes their responses, perhaps only temporarily.

Thus, it seems vitally important that researchers, in attempting to best represent the relationships in sample data, make diagnostic procedures aimed at detecting and assessing the impact of influential observations an integral part of regression analysis. Organizational research is clearly subject to a variety of situations that may bias data. By combining back end diagnostic steps with up front procedures to minimize bias, the best results are likely to emerge.

References

Bates, R. A. (1997). The impact of training content validity, organizational commitment, learning, performance utility, and transfer climate on transfer of training in an industrial setting. Unpublished doctoral dissertation, Louisiana State University.

Bates, R. A., Holton, E. F. III, Seyler, D. (1997). Factors affecting transfer of training in an industrial setting. In E. F. Holton III (Ed.), Academy of Human Resource Development 1997 Conference Proceedings.

Becker, T. E., Klimoski, R. J. (1989). A field study of the relationship between the organizational feedback environment and performance. Personnel Psychology, 42, 343-358.

Belsley, D. A., Kuh, E., Welsch, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. New York: Wiley.

Cardy, R. L., & Dobbins, G. H. (1994). Performance appraisal: Alternative perspectives. Cincinnati, OH: South-Western Publishing Co.

Cleveland, J. N., & Shore, L. M. (1992). Self- and supervisory perspectives on age and work attitudes and performance. Journal of Applied Psychology, 77(4), 469-484.

Darlington, R. B. (1990). Regression and linear models. New York: McGraw-Hill Publishing Company.

Ford, J. K., Quinones, M., Sego, D., & Sorra, J. S. (1992). Factors affecting the opportunity to use trained skills on the job. Personnel Psychology, 45, 511-527.

Freund, R. J., & Wilson, W. J. (1993). Statistical methods. San Diego: Academic Press, Inc.

Gielen, E. W. M. (1995). Transfer of training in a corporate setting. Unpublished doctoral dissertation, University of Twente, The Netherlands.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). Multivariate data analysis (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Hocking, R. R. (1976). The analysis and selection of variables in a linear regression. Biometrics, 32, 1-50.

Holton, E. F., & Russell, C. J. (1996). A longitudinal examination of socialization readiness, processes, outcomes, and turnover. Manuscript submitted for publication. Louisiana State University.

Peixoto, J. L., & LaMotte, L. R. (1989). Simultaneous identification of outliers and predictors using variable selection techniques. Journal of Statistical Planning and Inference, 23, 327-343.

Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer behavior and positive transfer of training. Human Resource Development Quarterly, 4(4), 377-390.

Smith-Jentsch, K. A., Jentsch, F. G., Payne, S. C., & Salas, E. (1996). Can pretraining experiences explain individual difference in learning? Journal of Applied Psychology, 81, 110-116.

Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. Journal of Applied Psychology, 76(6), 759-769.

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. Personnel Psychology, 48, 347-375.

Xiao, J. (1996). The relationship between organizational factors and the transfer of training in the electronics industry in Shenzhen, China. Human Resource Development Quarterly, 7(1), 55-86.